

5-25-2018

Analysis of Avian Diversity and Richness in Relation to Urban Park Size Using eBird

Jordan J. Tippetts
Portland State University

Follow this and additional works at: <https://pdxscholar.library.pdx.edu/honorstheses>

Let us know how access to this document benefits you.

Recommended Citation

Tippetts, Jordan J., "Analysis of Avian Diversity and Richness in Relation to Urban Park Size Using eBird" (2018). *University Honors Theses*. Paper 599.
<https://doi.org/10.15760/honors.609>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in University Honors Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Analysis of avian diversity and richness in relation to urban park size using eBird

JORDAN TIPPETTS

Abstract

I used data collected from eBird for 15 urban parks in Portland, Oregon to determine the effect of park area on species richness and diversity. Two data sets were extracted for each park; a full data set with the entire number of observations; and a reduced data set with 9 observations. Data sets were compared to determine if the same results and conclusions could be made with lower number of observations as compared to a much larger sample. In both data sets, species richness and diversity increased with park area. Although there was no significant difference in species diversity between data sets, richness was greater with more observations, indicating that more observations are needed to accurately calculate richness.

Introduction

Urbanization, the shift in human population from rural to urban areas, is steadily increasing throughout the world. As urbanization increases and becomes more prevalent, it becomes essential to determine the effects of urbanization on species and the environment because natural areas unavoidably decline as urbanization increases. Moreover, major urban centers are typically built near naturally biotically rich areas (e.g., confluence of major rivers) and thus urbanization has particularly severe negative consequences for native biota. One common approach to studying plant and animal diversity in urban landscapes is to apply MacArthur and Wilson's (1967) model of island biology which describes how area and isolation are thought to interact in determining the number of species found on habitat islands. It proposes that larger islands tend to have greater species richness than smaller islands, while islands that are more isolated tend to have lower species richness. Two of the main explanations given for this is that large islands have the greater ability to support larger populations that are less prone to extinction and larger islands have the capability to support a greater diversity of habitats.

Urban parks themselves can be viewed as habitat islands surround by an ‘urban ocean’ similar to, but different than, true islands. Indeed, one of the best predictors of avian diversity and richness is park size. Previous research has found that as urban patch size increases, so does avian diversity and richness (Kang et al., 2015) (Donnelly and Marzluff, 2004).

A relatively new resource being used to collect data is eBird. eBird, developed in 2002 by the Cornell Lab of Ornithology and the National Audubon Society, takes advantage of the enormous birding community to provide data on the distribution and abundance of birds throughout the world and through time. In addition, it also provides researchers with additional information including the duration of the observation, number of observers, distance traveled, and the observation protocol (traveling, stationary, or incidental). eBird is one of the largest citizen science project in the world, and each day tens of thousands of individuals submit observations to the database. Since launching, data input has increased exponentially (30-40% annually) from less than 1 million observations to over 140 million observations by mid-year 2013 (Sullivan et al., 2014). Likewise, the number of researchers, and government agencies using the data has increased steadily, and by 2013, more than 90 peer reviewed articles used eBird data or studied different aspects of the eBird project (Sullivan et al., 2014). These topics range from descriptions of species’ distribution, timing of migration, evolution, responses to climate change, and biogeography.

A long-standing concern with the utilization of all citizen science programs is data quality (Sauer, Peterjohn, & Link, 1994), and with eBird, the primary concern is misidentification of bird species (Kelling et al., 2012). However, eBird has taken many steps

towards improving data quality including the implementation of a two-part verification process (Sullivan et al., 2009). The first step is automatic and implemented at the point of data entry. When an individual begins an observation, eBird automatically generates a list of species known to be present based on their location and date. Species and their abundances are then recorded by the observer(s). If observations of species or abundances are not common for the location or date, they are automatically flagged for review. This then gives the user the opportunity to correct any accidental errors, reconsider a species' identity, or if confirmed, provide any supporting evidence. Once confirmed, if the observation is still flagged, it is sent to a regional expert to determine the validity of the observation based on the evidence provided. Confirmation of the utility of e-bird data for addressing science-based questions must be made through comparisons of professional and citizen science data set from the same area. Indeed, after correcting for differences in intensity of sampling, Callaghan and Gawlik (2015) showed that species diversity and richness at the site of a small-scale estuary restoration project located in south eastern Florida did not differ significantly.

A question of importance relating to data collection and analysis is what the minimum sample size needed to produce reliable and accurate results. Under-sampling will lead to error while oversampling is a waste of resources that could otherwise be used elsewhere. eBird shares this same concern of needing to know how many checklists constitutes a legitimate sample. A single number is no doubt not applicable to all studies because of large differences in true species richness and abundance, and habitats and/or geographic regions of low natural species richness will likely take fewer samples to provide repeatable results. Only through comparisons made by multiple individuals at numerous locations will a better understanding of

minimum sampling needs be known. However, in highly urbanized areas, environments tend to be similar and thus a generalized prediction for the number of observations needed may be predictable. In a recent paper by Callaghan, Lyons, Martin, Major, and Kingsford (2017), they obtained data from eBird on the most populated areas in the United states and Canada and determined the minimum number of checklists needed to fully sample an urban area using eBird. They found that the average number of observations needed to obtain accurate estimates of species richness and species diversity was 210 and 33, respectively. However, when species that were present on 5% or fewer of the checklists were removed the average number of checklists decreased to 17 and 9, respectively.

In this report, I used observational data collected from eBird of 15 urban parks in Portland, OR (USA) to determine the effect of park area on species richness and diversity. I then compared two data sets; a full data set and a reduced data set to determine if they generated the same results such that the smaller sample would lead to the same conclusions as the much larger sample. I predicted that (1) Larger parks would have higher richness and diversity than smaller parks, (2) diversity would not differ significantly between the full and reduced data sets, (3) richness would differ between the two data sets, and that (4) parks located in direct contact with a major river will have higher richness and diversity.

Methods

Study Sites

Data was extracted from eBird for fifteen urban parks in Portland, Oregon for observations made during the 2016-2017 breed season. For each park, the following criteria were used to select checklists and obtain data for analysis: (1) checklists had to be complete, meaning

that checklists that did not record all species or marked an “X” for species abundance were removed; (2) as not all parks had bodies of water near them, my analysis was restricted to terrestrial species; (3) checklists were removed if duration of visits was shorter than 30 minutes and longer than 200 minutes; (4) checklists with group identifiers, signaling duplicate checklists from groups, were excluded; and (5) checklists were only included if protocol was labeled as stationary or traveling (see Sullivan et al. 2014). In addition, species present on 5% or less of the checklists for each park were removed to decrease the total number of checklists needed to obtain accurate measurements of diversity (9 observations needed) and richness (Callaghan et al., 2017). Data sets were compared using species richness (S), the Shannon Diversity Index (H'), and the number of equally common species that would produce same diversity as H' ($N1$). Species richness is the total number of species observed. The Shannon diversity index differs from richness in that it incorporates both number of species and their relative abundance. Thus, the most diverse communities are those with the greatest number of species and with relatively equal numbers of individuals of all species (i.e., high equitability).

Table 1. Parks from Portland, OR, used in the analysis, along with their area (ha), whether they were in direct contact with a large river, and total number of counts recorded in e-bird (number surveys).			
Park	Area (ha)	Major River	Number surveys
Albert Kelly Park	4.888607	No	24
Audubon	60.7029	No	27
Broughton Beach	0.46575	Yes	55
Crystal Spring Rhododendron Garden	3.84047	No	26
Gabrel Park	36.68479	No	151
Mt. Tabor	77.22218	No	57
Oaks Bottom	57.14976	Yes	49

Pittock Mansion	18.61556	No	57
Powell Butte	247.6597	No	42
Reed Lake	8.91	No	177
Sandy River Delta	566.5604	Yes	97
Tryon Creek	261.0225	No	34
Vanport Wetlands	36.62408	No	179
Wapto Access Greenway SP	6.07029	Yes	141
Whitaker Ponds Nature Park	10.01598	No	34

Park area was not normally distributed and therefore to conform to requirements of parametric analyses, I \log_{10} transformed park area for all statistical analyses. To evaluate the relationship between park area and species richness and diversity, two analytical tests were used. A least squares linear regression was used for species diversity and N1 because both variables exhibited a continuous range of variation. Strictly speaking, least squares linear regression is inappropriate for response variables that are counts such as number of species. For the latter, the preferred method is Poisson regression. The strength of the relationship between two variables in least squares linear regression is indicated by the coefficient of determination (r^2) which varies from 0 (no relationship) to 1.0 (perfect 1:1 relationship). Poisson regression does not formally calculate an r^2 value, but a “pseudo r^2 ” value was calculated to determine the percent of variation in species richness that park area accounted for. All statistical analysis was completed using the software STATISTIX.

To test whether there was a difference in diversity, N1, and richness between limited samples as compared to much larger samples, two data sets were collected for each park; a full data set, and a reduced data set. The full data set included all observations recorded over the

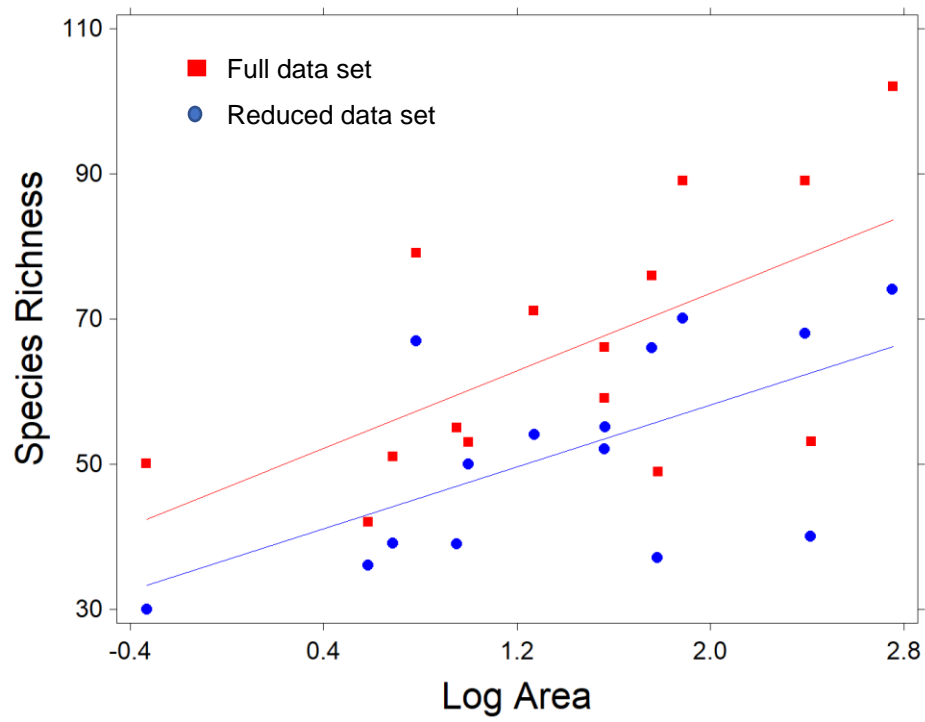
course of the year (2016-2017; Table 1). Diversity and richness measurements in the full data set were presumed to be comparable based on assumptions from accumulation curves for species richness and diversity. The reduced data set included 9 randomly chosen observations. To determine whether the reduced data set was different in terms of diversity, N1, and richness a two-sample pooled t-test was used. Analyses were repeated for each data set and compared to determine if results were the same.

Results

Area

A total of 1150 surveys were collected for the full data set with the number of observations at each park averaging $76.7 (\pm 14.7 \text{ SE})$ but ranging from 25 to 179 (Table 1). Poisson regression demonstrated that richness increased significantly with area (coefficient $[\beta] = 0.208 \pm 0.041 \text{ SE}$, $P < 0.001$; Fig. 1A), while least squares regression showed that both Shannon-Wiener diversity ($\beta = 0.233 \pm 0.077 \text{ SE}$, $P = 0.010$; (Fig. 1B) and N1 ($\beta = 5.65 \pm 1.923 \text{ SE}$, $P = 0.012$) increased with park area. Area alone accounted for 38.9%, 41.2%, and 39.9% of the variation in richness, Shannon-Wiener diversity and N1, respectively. Proximity to water contributed significantly to variation in species richness ($\beta = 0.247 \pm 0.069 \text{ SE}$, $P < 0.001$; Fig. 2), but not to Shannon-Wiener diversity ($\beta = 0.202 \pm 0.132 \text{ SE}$, $P = 0.152$) or N1 ($\beta = 6.21 \pm 3.123 \text{ SE}$, $P = 0.070$). Addition of proximity to water increased the explained variation in species richness, Shannon-Wiener diversity and N1 to 57.7%, 50.8%, and 54.8%, respectively.

(A)



(B)

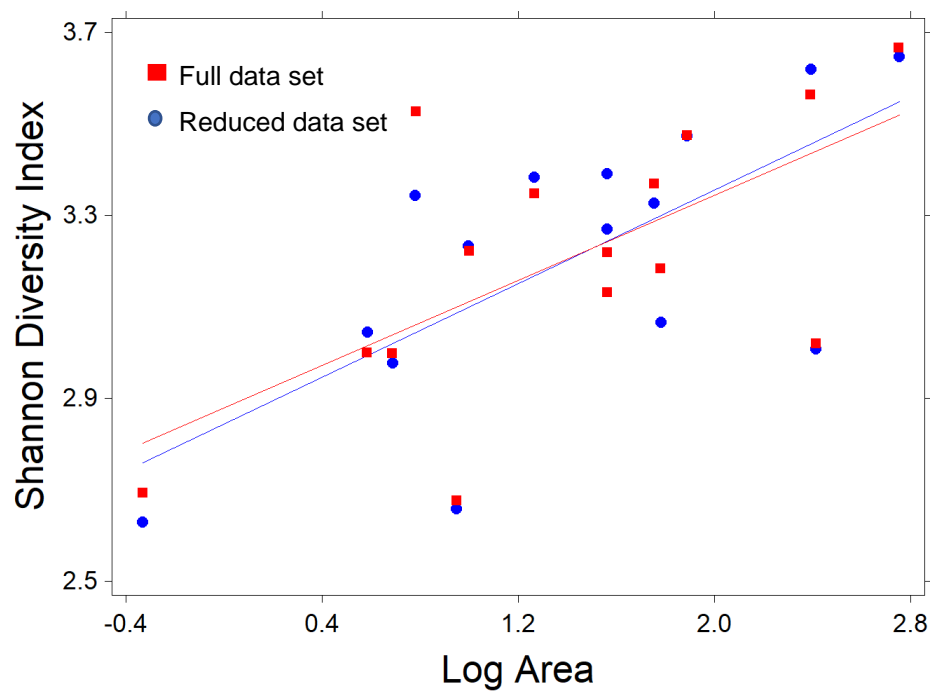


Figure 1: Species richness (A) and species diversity (b) versus park area for 15 urban parks in Portland, Oregon.

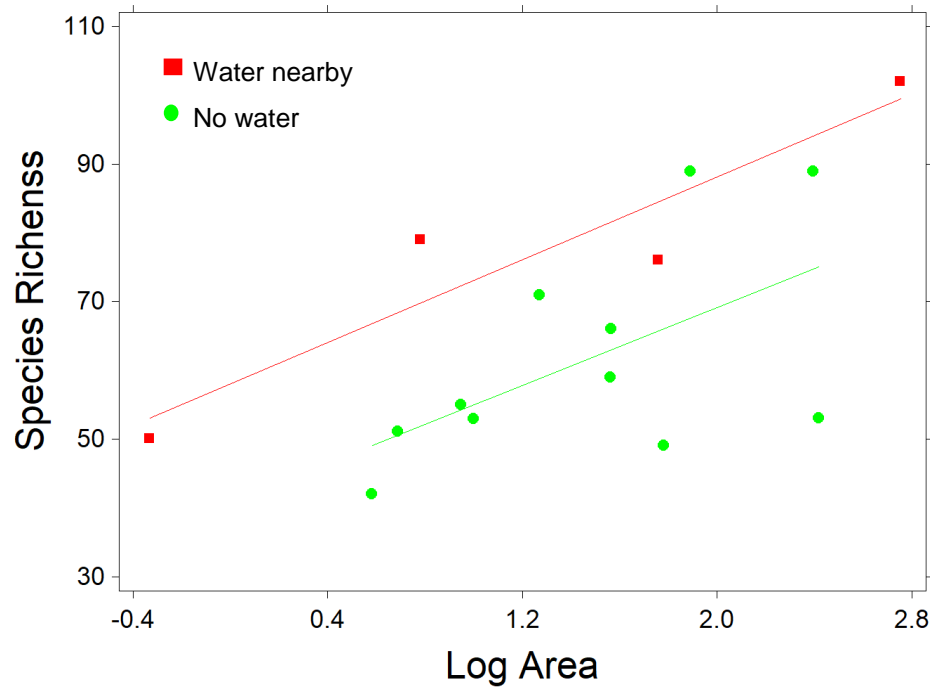


Figure 2: Difference between parks with and without direct contact to a large river in species richness versus park area for 15 urban parks in Portland, Oregon.

Comparison of full and reduced data sets

The restriction of the original number of surveys to nine represented a reduction to just 18.5% of the original number ($\pm 2.88\%$ SE, range = 5.0% to 37.5%). As in the larger data set, species richness ($\beta = 0.210 \pm 0.047$ SE, $P < 0.001$), Shannon-Wiener diversity ($\beta = 0.256 \pm 0.076$ SE, $P = 0.005$), and N1 ($\beta = 6.21 \pm 1.806$ SE, $P = 0.004$) all increased with park area. Respective explained variation for the three variables were 36.1%, 46.6%, and 47.6%. Proximity to water was again associated with an increase in species richness ($\beta = 0.215 \pm 0.078$ SE, $P = 0.007$), but not Shannon-Wiener diversity ($\beta = 0.102 \pm 0.139$ SE, $P = 0.476$) or N1 ($\beta = 3.35 \pm 3.242$ SE, $P = 0.321$). The regression line describing the relationship between species richness and park area appeared to be higher in the full sample compared to the sample of nine surveys (Fig. 1A) and analysis of covariance confirmed that the full sample detected more species when

park area was accounted for ($F = 8.11$, $df = 1$, 27 , $P = 0.008$). On average, 13.8 more species (± 1.68 SE) were detected by the full sample than the reduced sample of nine surveys (paired t-test = 8.23, $df = 14$, $P < 0.001$). By contrast, a similar analysis of covariance for species diversity (Shannon-Wiener diversity) revealed that the full and reduced samples produced virtually identical estimates of diversity ($F = 0.00$, $df = 1$, 27 , $P = 0.9791$; Fig. 1B), and a paired t-test indicated that the average difference in N1 between the two data sets (-0.01 ± 0.0628 SE) was negligible ($t = 0.02$, $df = 14$, $P = 0.988$).

Species-area relationships

Standard island biogeographic approaches to the study of variation in species richness relate richness to area as a \log_{10} - \log_{10} relationship in which richness (S) is expressed as a power of area ($S = cA^z$; MacArthur and Wilson 1967). “ z ” represents the slope and “ c ” is a constant reflecting the y-intercept. The power equation relating richness to area for the full data set is

$$S = 48.31A^{0.084},$$

While that for the reduced data set is,

$$S = 36.55A^{0.094}.$$

The difference in the number of species predicted by the two equations for habitat islands of 5 ha, 15 ha, 25 ha, 50 ha and 100 ha averaged 13.8 (± 0.36 SE), which is identical to the 13.8 more species detected on average by the larger sample (see above).

Discussion

Previous research from both true oceanic islands (e.g., Murphy et al. 2004) and islands of habitat in terrestrial environments (Donnelly et al. 2004, Kang et al. 2015) have shown that area has a strong influence on number of species present. My results thus conform well to previously published works. However, it is important to also acknowledge that many other factors, most of which were not accounted for, could influence species richness. For instance, richness was higher in parks located near bodies of water, which likely represents an increase in habitat diversity that would attract additional species. Other factors that could affect richness in urban parks that were not accounted for include number of habitat types, degree of isolation from other suitable habitat, and degree of urbanization in the immediate area surrounding a park. All have been shown to impact species richness and diversity in parks and habitat islands. Only about 35-40% of the variation in species richness was accounted for by park area, and one reason for the relatively low explained variation in richness is that many of the parks were located close to other parks and/or had large areas of surrounding forest in the vicinity. Thus, it is not apparent in this study how much park size specifically contributes to diversity and richness and what is just a result of other contributing variables.

Species richness was significantly higher in the full data set compared to the reduced data set. Callaghan (2017) reported a similar finding and he found that at least 17 surveys were needed to accurately predict species richness in urban areas. By contrast, Shannon-Wiener diversity and N1 did not differ significantly between the full and reduced data sets despite the fact that the reduced number of surveys (nine) represented less than 20% of the average number of surveys contributing to the full data set. This suggests that nine observations were enough to

accurately calculate species diversity and N1 of urban parks in Portland, Oregon which is consistent with past research in urban areas using eBird (Callaghan, 2017). However, given that my research only tested the accuracy of nine observations, it is possible that even fewer observation may be needed to accurately predict species diversity. Why more surveys are needed to accurately measure species richness using e-bird is not entirely clear, but a possible contributing factor is that eBirders continue to find more species for local patches over an extended time period (Callaghan, 2017). These additional species are typically represented by one or just a few individuals and are thus “rare”. Rare species contribute relatively little to measures of overall species diversity that take evenness of abundance (i.e., equitability) into account, and this likely explains why species richness, but not species diversity, was affected by number of surveys.

Although species richness increased with park area, the effect of park area on richness for both the full and reduced data set was relatively low ($z=0.084$ and $z=0.094$) when compared to true oceanic islands which z -values typically range from 0.2 – 0.4 (MacArthur and Wilson, 1967). One possible explanation for this is that the ocean that separates true islands from each other presents a bigger barrier to dispersal for terrestrial bird species than the space in-between urban parks which often contains vegetation, food, and fresh water. Thus, the effect of park area and isolation is likely very dependent on what is in the space in-between urban parks. Even though MacArthur and Wilson’s (1967) model of island biology provides a good starting point for studying urban landscapes, there are clearly many more variables that need to be accounted for in urban areas.

Future research should include a more in-depth analysis of other variables that have been found to influence park diversity and richness in urban parks such as plant species richness, elevation, age of park, and fragmentation. In addition, when using eBird, a greater variability in the number of observations should be tested to acquire the minimum number of observations needed to accurately calculate species richness and diversity.

Literature Cited

Callaghan, C. T., & Gawlik, D. E. (2015). Efficacy of eBird data as an aid in conservation planning and monitoring. *Journal of Field Ornithology*, 86(4), 298-304.

Callaghan, C., Lyons, M., Martin, J., Major, R., & Kingsford, R. (2017). Assessing the reliability of avian biodiversity measures of urban greenspaces using eBird citizen science data. *Avian Conservation and Ecology*, 12(2).

Donnelly, R., & Marzluff, J. M. (2004). Importance of reserve size and landscape context to urban bird conservation. *Conservation Biology*, 18(3), 733-745.

Kang, W., Minor, E. S., Park, C. R., & Lee, D. (2015). Effects of habitat structure, human disturbance, and habitat connectivity on urban forest bird communities. *Urban ecosystems*, 18(3), 857-870.

Kelling, S., Gerbracht, J., Fink, D., Lagoze, C., Wong, W. K., Yu, J., ... & Gomes, C. (2012). A human/computer learning network to improve biodiversity conservation and research. *AI magazine*, 34(1), 10.

MacArthur, R. H., & Wilson, E. O. (2016). *The theory of island biogeography*. Princeton University Press, Princeton, NJ.

Sauer, J. R., Peterjohn, B. G., & Link, W. A. (1994). Observer differences in the North American breeding bird survey. *The Auk*, 50-62.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282-2292.

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., ... & Fink, D. (2014). The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, 169, 31-40.